접수번호	※작성하지 않음

「2025년 통계데이터 활용대회」 데이터분석 보고서

제 목	스코어링 분석을 통한 ‘혼행족’ 맞춤 관광지 발굴

신 청 자 명	소속/직위	한양대학교 등	성 명	안태영, 김채원, 구세연
	휴대전화		전자우편	
제출일	2025.07.02.			

스코어링 분석을 통한 ‘혼행족’ 맞춤 관광지 발굴

1. 배경

□ 주제 선정 배경

혼자 하는 여행을 뜻하는 ‘혼행’은 1인 가구 증가와 사회적 거리 두기에 따른 사회적 배경에 따라 새로운 여행 행태로 등장했다. ‘혼행족’은 COVID-19 이전인 2018년에는 2.5% 정도에 불과했지만 COVID-19 이후로 꾸준히 3% 이상을 유지하고 있으며 앞으로도 증가할 여지가 높다.

한편 2024년 여행지별 1인 평균 국내 여행 횟수는 경기(1.27회), 강원(0.71회) 등의 순으로 높지만(문화체육관광부, 2024), 이는 예년과 비슷한 양상이다. 관광객들에게는 기존의 익숙한 관광지 외에도 새로운 관광지가 요구되는 상황이며, 더 많은 관광객 유치가 필요한 지역들의 기반 마련을 위한 지표가 필요한 상황이라고 판단했다.

따라서 스코어링 분석을 통해 혼행족의 성향과 선호에 맞는 맞춤형 관광지를 발굴하고, 나아가 지역 관광 활성화와 연결시킬 수 있는 데이터 분석 모델을 만들어보고자 한다.

□ 분석 필요성 및 전략

1. 분석 필요성

본격적인 연구를 진행하기에 앞서 기존에 비슷한 주제로 논의된 연구를 몇 가지 분석하였다. 이순학(2017)은 여행 행태 만족도에 기반하여 1인 여행객의 여행 행태를 분석하고 1인 여행객 확대를 위한 정책적 시사점을 도출하고자 했다. 한혜원(2024)에서는 로컬 여행 진흥을 위한 앱 서비스 필요성에 대해 논의하며, 지역 주민의 참여를 기반으로 한 콘텐츠와 인터페이스 디자인을 제안한다.

본고에서는 앞선 연구들에서 진행되지 않았던, 관광 산업 측면에서의 국내 지역에 대한 분석을 통해 ‘혼행족’을 위한 맞춤 관광지를 발굴해 보고자 한다.

2. 분석 전략

본고에서는 ‘혼행족’들이 여행할 때 고려할 요소(이하 ‘여행 요소’)로 4가지를 선정하였으며, 그 요소는 접근성, 편의시설, 볼거리, 음식이다. 주요 도시인 서울, 광역시 6개, 세종, 제주를 제외한 총 151개 국내 지역을 ‘분석 지역’으로 선정하였으며, 해당 지역들의 방문율과 여행 요소 4가지의 상관관계를 데이터 분석을 통해 도출한다. 데이터 분석 결과를 토대로 관광에 영향을 미치는 여행 요소가 무엇인지 선별하고, 각 여행 요소에 점수를 부여하는 스코어링 분석을 시행한다.

2. 데이터 분석

□ 데이터 선정

상술한 바와 같이 본고에서 지정한 여행 요소는 접근성, 편의시설, 볼거리, 음식이며, 따라서 활용 데이터 또한 여행 요소와 관련된 것으로 선정하였다. 단, 음식 카테고리 중 '소비자당 매출'은 내국인국내카드소비(BC카드) 데이터 중 음식 관련된 사용 내역만 사용한다.

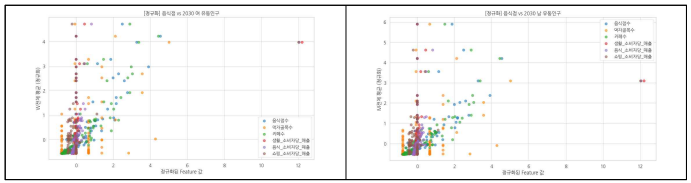
여행 요소	여행 요소_하위 분류		원천 데이터_데이터명	원천 데이터_출처
접근성	기차역		국가철도공단_철도역 정보	공공데이터포털
	터미널		전국 시외버스 터미널 안내	전국 고속/시외 버스 운행정보
	렌터카 업소		전국렌터카업체정보표준데이터	공공데이터포털
	택시		시군별 면허대수 현황	택시 감차보상재원 관리기관
	시내버스(노선 개수)		버스노선별 운행지역 정보	산림빅데이터거래소
	시내버스 (평균 배차 간격)			
	공항		한국공항공사_공항 위치정보	공공데이터포털
편의 시설	숙박 업소		관광숙박업	LOCALDATA
			숙박업	
	공용화장실		전국공중화장실표준데이터	
	편의점		행정안전부_생활안전지도 편의점	생활안전정보
	쇼핑센터		국내여행 소비 역세권지도	문화 빅데이터 플랫폼
볼거리	레저		레저사업장 일반 정보	문화 빅데이터 플랫폼
	전통시장		소상공인시장진흥공단_전통시장현황	공공데이터포털
	박물관미술관		전국박물관미술관정보표준데이터	
	산책로		내 주변 산책로 데이터	문화 빅데이터 플랫폼
	등산로		경기도_숲길 현황, korea sanha	공공데이터포털
	해수욕장		해수욕장 개폐장일정 정보	
	문화유적지		전국향토문화유적표준데이터	
음식	음식점		지역별 음식점 소비기반 트렌드	문화 빅데이터 플랫폼
	먹자골목		먹자골목 정보	
	카페		지역별 루프탑 호텔라운지 카페 식당 수영장 바등 공간 정보 및 글램핑 정보	
	소비자 당 매출	생활	내국인국내카드소비(BC카드)	통계데이터센터 _민간자료
		음식		
		쇼핑		

[표 1] 데이터 명세서

해당 데이터들은 데이터 분석에 앞서, 필요 지역의 데이터만 추출하거나 중복 내용은 삭제하는 등의 전처리 과정을 거친 후 분석에 사용했다.

□ 데이터 분석

1. 탐색적 데이터 분석 - EDA(Exploratory Data Analysis)



[표 2] 네 개의 카테고리 유동인구별 분포도

[표 2]는 네 개의 카테고리인 음식, 볼거리, 이동수단, 편의시설을 N(0, 1)로 정규화하여 유동인구에 대해 남녀로 나누어 도표로 나타내었다.

2. 변수 필터링 및 다항식 전개

본 과정부터는 네 개의 카테고리에 같은 방법론을 적용하였으므로 음식 카테고리를 예시로 들어 결과를 나타내었다.

2.1 Polynomial Features

변수 간의 관계성과 비선형성을 탐색하기 위해 차수를 2로 설정하여 독립변수를 증가시켰다. [그림 1]은 음식 카테고리를 차수를 높여 N(0, 1)로 정규화한 결과다.

2.2. Lasso CV

앞서 채택한 변수를 회귀 모델에서 중요하지 않은 변수의 가중치를 0으로 근사하여 모델을 간결하게 해주는 회귀 방법으로, 각 변수의 영향력을 파악하여 핵심 변수를 선별하기 위하여 사용하였다. [그림 2]는 Lasso CV를 적용한 결과다.

2.3 VIF (Variance Inflation Factor)

특정 변수가 다른 변수들에 의해 얼마나 잘 설명되는지를 수치화한 값으로 VIF 값이 크면 클수록 해당 변수는 다른 변수들과 강한 상관관계가 있다는 뜻으로 독립변수로 역할을 할 수 없다고 해석하여 해당 변수의 VIF가 10보다 제거하는 방법을 사용하였다. [그림 3]은 VIF < 10을 적용한 결과다.

2.4 OLS(Ordinary Least Squares)

앞서 소개한 방법론을 거친 변수 중에서 회귀 모델링을 적용했을 때 통계적으로 유의미한 값을 알아보기 위해 p-value < 0.1(90% 신뢰구간 안)에만 존재하는 변수들로 간추려 회귀 모델링을 진행하였다. [그림 4]는 OLS를 통해 p-value < 0.1을 적용한 결과다.

	음식점수	먹자골목수	카페수	생활_소비자당_매출	음식_소비자당_매출	쇼핑_소비자당_매출	음식점수^2	음식점수 먹자골목수	음식점수 카페수	음식점수 소비자당_매
0	0.519242	0.661888	0.820527	0.012430	0.412740	-0.296963	0.269613	0.343680	0.426052	0.006
1	-0.653430	-0.786591	-0.643307	-0.138854	-0.184673	0.000000	0.426971	0.513982	0.420356	0.090
2	-0.301504	-0.786591	-0.352841	-0.116982	-0.161103	0.000000	0.090905	0.237161	0.106383	0.035

[그림 1]

	음식점수	카페수^2	생활_소비자당_매출	음식_소비자당_매출	음식_소비자당_매출^2	쇼핑_소비자당_매출^2
0	0.519242	0.673264		0.005130	0.170354	0.088187
1	-0.653430	0.413843		0.025643	0.034104	0.000000
2	-0.301504	0.124497		0.018846	0.025954	0.000000

[그림 2]

	음식점수	카페수^2	쇼핑_소비자당_매출^2
0	0.519242	0.673264	0.088187
1	-0.653430	0.413843	0.000000
2	-0.301504	0.124497	0.000000

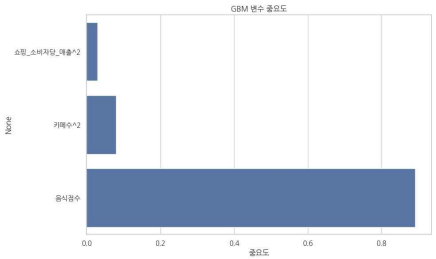
[그림 3]

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.01e+05	2.52e+04	11.944	0.000	2.51e+05	3.51e+05
음식점수	4.663e+05	3.5e+04	13.334	0.000	3.97e+05	5.35e+05
카페수^2	4.036e+04	1.49e+04	2.711	0.008	1.09e+04	6.98e+04
쇼핑_소비자당_매출^2	3578.5264	1879.255	1.904	0.059	-135.320	7292.372

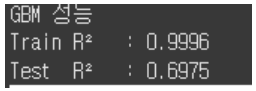
[그림 4]

2.5. 모델링 – GBR(Gradient Boosting Regressor)

GBR, XGBoost, LightGBM, OLS 네 가지 알고리즘을 적용해본 결과 가장 데이터를 잘 표현한 모델은 GBM 모델이었다. GBM 모델은 데이터가 선형적인 구조라고 가정을 할 수 없을 때 사용하는 방법론으로 비선형성과 변수 간 상호작용까지 반영한 예측 회귀 예측 모델이다. 하이퍼파라미터를 통해 파인튜닝을 한 결과 이상치에 민감하지 않은 'huber' loss를 사용하였고, alpha=0.9, n_estimators=200, learning_rate=0.1, max_depth=4, subsample=0.8을 사용하여 모델링을 진행하였다. [그림 6]은 GBM 모델의 R^2 값이고 [그림 5]는 GBM 모델의 변수 중요도를 도표로 나타내었다.



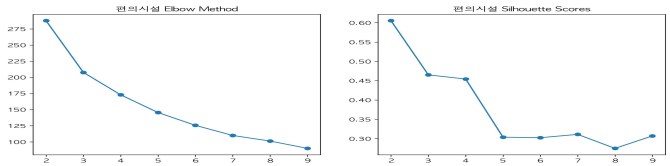
[그림 5] GBM R^2 값 및 변수 중요도 결과



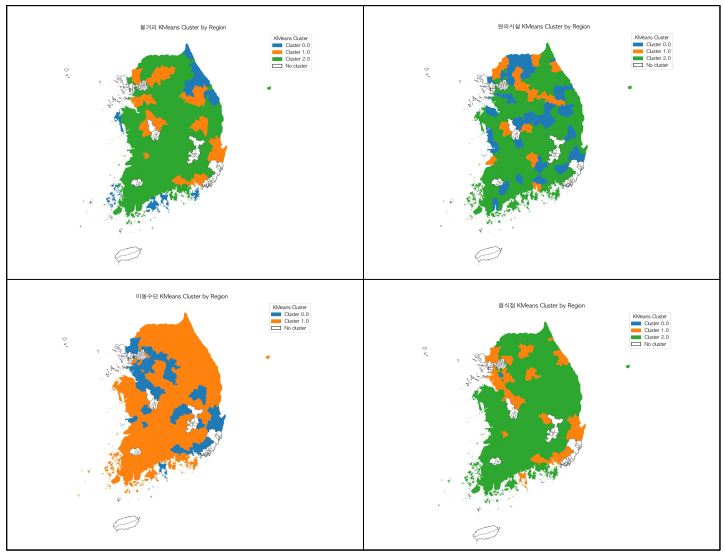
[그림 6] GBM R^2 값

□ 분석 결과 및 해석

본 연구는 각 지역의 볼거리, 편의시설, 이동수단, 음식점 features를 기반으로 유사한 특성을 지닌 지역들을 군집화하였다. 각 요소의 features 정규화, KMeans clustering, PCA, 시각화 순으로 분석을 진행하였다. [그림 7]은 편의시설의 엘보우 기법과 실루엣 계수를 나타낸 것으로 K=3에서 엘보우 지점이 포착되고, 실루엣 도표상 클러스터 폭도 균등하게 형성된 것을 확인할 수 있다. 볼거리, 이동수단, 음식점의 군집 수 설정에도 같은 방법론을 적용하였다.



[그림 7] 편의시설 엘보우 기법과 실루엣 계수



[그림 8] 지표별 군집 지도 시각화

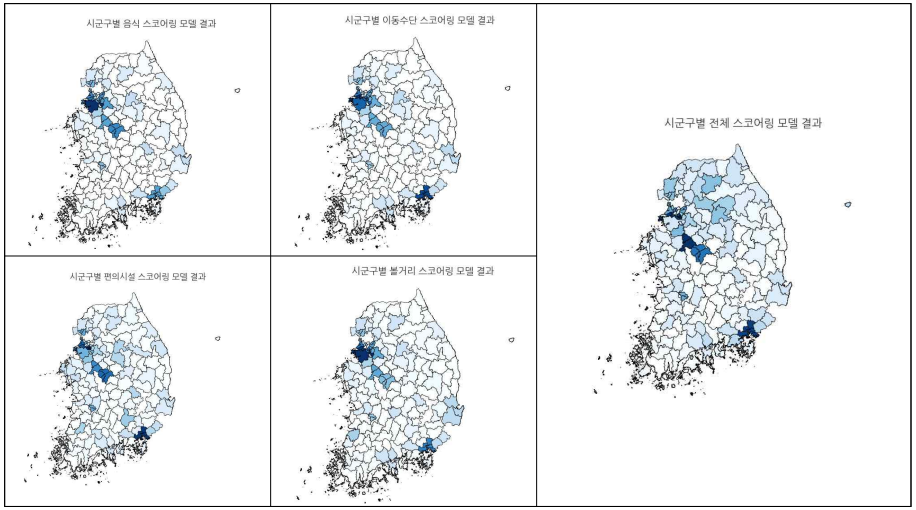
분석 결과, 볼거리 지표는 Features의 정규화 수치에 따라 세 개의 군집으로 구분된다. cluster 0은 강원도와 같이 자연경관 중심의 관광지, cluster 1은 공주시, 경주시 등 문화 자원 중심의 관광지, cluster 2는 전반적으로 관광 비중이 낮은 지역군으로 해석된다. 따라서 볼거리 인프라가 부족한 지역이 많다.

편의시설 지표는 features의 정규화 수치 순으로 대도시형 지역군, 중소형 지역군, 비도시형 지역군으로 세 개의 군집이 추출되었다. [그림 8]에 따르면 경기도와 충청북도 일부 지역 외에 편의시설 인프라가 부족한 지역들이 많으며 해당 지역의 관광 수요가 낮다고 분석하였다.

이동수단 지표를 통해 경기도, 충청남도 일부, 경상남도 일부, 경상북도 일부 지역만이 우수 지역으로 확인되었고, 교통 인프라 열위 지역은 혼행 관광지로 잘 선정되지 않는다고 해석할 수 있다.

음식점 지표는 외식 인프라에 따라 세 개의 군집으로 나뉘었으며, 수원시만이 높은 수준으로 나왔고, 그다음으로 경주시, 포항시, 강릉시 등이 중간 수준의 인프라를 보유하고 있다.

위 결과를 토대로 부족한 지표의 features를 보완하는 지역별 관광 전략을 수립할 수 있을 것이다.



[그림 9] 각 여행 요소별 스코어링 분석 결과

[그림 9]는 네 개의 카테고리에 대해서 스코어링 모델을 시군구별로 시각화한 결과다. 광역시는 본 연구에서 제외된 지역으로 공백으로 표현하였다. 시각화 결과, 수도권 지역과 경남권이 높은 점수를 보이는 것으로 나타났고, 해안 지역이 다음으로 높은 점수를 받는 경향성이 나타났다.

3. 분석 활용 전략

□ 기대효과

도출된 데이터 내용으로 혼행족에게 로컬 여행 지역을 추천할 수 있으며, 나아가 관광 사업 육성이 필요한 지역에는 실현 가능하며 구체적인 제언 방향을 제시할 수 있는 계기가 될 것이다. 구체적인 사용 방향은 다음과 같다.

□ 방향 제시

1. 로컬 여행 추천 앱 활용

해당 스코어링 모델을 바탕으로 어플리케이션을 통해 사용자의 수요에 맞춰 로컬 여행 지역을 추천할 수 있다. 대중적으로 알려진 관광 지역 외에도 여행 요소에 따라 더 다양한 지역이 사용자에게 노출될 수 있을 것이다.

2. 로컬 여행 지역 발전 방향 제언

로컬 여행의 중점은 지역 특성은 유지하면서도 표준화된 서비스 품질과 관관객 응대 방안 마련(진보라, 2022)이 필요하다는 점에 있다. 관광객들에게 어필할 만한 요소가 많은 지역임에도 불구하고 관광객 유치가 어려웠던 지역은 해당 분석 내용을 활용해 부족한 인프라 제고에 힘을 쓸 수 있을 것이다.

참고문헌

문화체육관광부(2019). **2018년 국민여행조사 보고서 -분석편-**.
문화체육관광부(2023). **2022년 국민여행조사 보고서 -분석편-**.
문화체육관광부(2024). **2023년 국민여행조사 보고서 -분석편-**.
문화체육관광부(2025). **2024년 국민여행조사 보고서 -분석편-**.
이순학(2017). **1인 여행객의 국내 여행 행태분석과 시사점**. 산업연구원.
진보라 외 3인(2022). **[관광] 관광 트렌드 분석 및 전망 2023-2025**. 한국문화관광연구원.
박진영, 이상재(2020). **사이코그래픽 시장세분화를 통한 대학생들의 국내여행 선호도에 관한 연구**. Tourism Research 45권 3호.
한혜원 외 3인(2024). **지역 경제 활성화를 위한 사용자 참여 기반 로컬여행 진흥 앱 서비스 제안: MZ세대를 대상으로**. 한국디지털콘텐츠학회논문지 제25권 제7호.